speech sound is not distorted.

In the case of "Adaptation by only distortion of spectral frequency", a single trained pattern is used and all training speakers are not categorized, similarly to the case of "No adaptation". While, speaker adaptation processor 205 distorts spectral frequency of a user's speech sound.

In the case of "Adaptation by pattern selection and distortion of spectral frequency" in coincidence with the present invention, trained patterns are generated in response to ages of speakers. Pattern by-characteristic selection unit 204 selects a trained pattern. Also, speaker adaptation processor 205 distorts spectral frequency of a user's speech sound.

Comparison of these methods results in effectiveness of "Adaptation by pattern selection and distortion of spectral frequency", for all age groups, namely 12 or lower, 13 through 64, and 65 or higher.

In the present embodiment, the stored trained patterns on pattern by-characteristic storage 203 are categorized according to ages of training speakers. However, the trained patterns may be categorized according to regions where the training speakers live or lived for the longest time, or mother tongues of the training speakers.

In the present embodiment, a microphone and a voice controller are integrated, and a control signal is sent to each controlled device. However, a microphone and a voice controller may be incorporated in each device.

Additionally, speech recognition apparatus 211 in coincidence with the present embodiment has pattern selection word file 209 which includes the pattern selection words as known words, as a part of word lexicon 208. However, without file 209, the most similar category can be selected in this way: the most similar speech sound elements and every speech sound of the first utterance are lined up, then the distances between these two lines are compared.

Exemplary embodiment 2

Fig. 6 shows a second exemplary embodiment. This embodiment differs from exemplary embodiment 1 in that a user can arbitrarily register a pattern selection word in pattern selection word file 609. The other points remain the same as those in exemplary embodiment 1.

For registering a new name for a device, a user speaks word "Register" as a first utterance. Word "Register" is previously stored as a recognized word in word lexicon 208. Speech recognition unit 606 is transferred into a

vocabulary registering mode in response to the utterance. Pattern by-characteristic selection unit 204 recognizes "Register" as a pattern selection word similarly to the recognition of the first utterance in embodiment 1. Simultaneously with the recognition, pattern selection unit 204 determines trained patterns, and speaker adaptation unit 205 determines a distortion coefficient of speech sound spectral frequency. The trained pattern and the distortion coefficient are used for later speech recognition. Speech recognition unit 606 performs, using the distortion coefficient, a spectral frequency distortion calculation for an LPC cepstral coefficient vector of a next new name spoken by the user, for example, "Lamp". Phonemes are arranged and fitted from the trained patterns to be selected so that phonemic inconsistency does not occur, thereby obtaining an acoustic unit arrangement corresponding to the utterance of "Lamp". Speech recognition unit 606 stores this pattern arrangement or a character string "Lamp" converted from the arrangement on word lexicon 608. The arrangement or the character string is set to be a new pattern selection word for a first utterance

The registered pattern selection word is used from now on for speech recognition similarly to the other pattern selection words. Even if a user speaks "Lamp_Turn off" instead of "Light_Turn off", for example, the user can obtain the same result. When a device control word is previously defined to every device, the newly registered pattern selection word must be related to a device control word.

In embodiment 2, pattern selection unit 204 determines trained patterns to be selected based on an utterance "Register". However, a previously defined typical trained pattern may be used for reducing total time of the registering process.

Exemplary embodiment 3

Fig. 7 shows a third exemplary embodiment. A process is added to exemplary embodiment 1. It is a process for resetting the trained pattern selected by a user's first utterance and a distortion coefficient of speech sound spectral frequency and for setting a user's next utterance to be a first utterance.

Reset signal generation unit 701 detects an output from speech recognition unit 206 to control signal output unit 207, and informs acoustic analysis unit 202 of the completion of the user's utterance for device control.

Acoustic analysis unit 202, when it receives this notification, resets the

receiving state.   Unit 202 moves to the following mode: unit 202 sets next input speech sound from sound input unit 201 to be a first utterance by the user, and supplies an LPC cepstral coefficient vector to pattern by-characteristic selection unit 204 and speaker adaptation processor 205.   The user always speaks a pattern selection word such as a device name and a device control word in a pair, thereby operating the device in high recognition accuracy.

Embodiment 3 uses the output from speech recognition unit 206 as timing for resetting acoustic analysis unit 202.   When speech recognition unit 711 makes a recognition error, this recognition error timing may be set to be a receipt timing of a reset instruction supplied from a key or the like. Additionally, the reset timing may be obtained when speech recognition unit 206 outputs nothing for a predetermined period.   In this case, a timer is disposed in reset signal generation unit 701.

Speech recognition unit 711 can discriminate with reset signal generation unit 701 whether or not an utterance is a first utterance or a subsequent utterance.   It is not required to always initially determine whether or not an utterance is a name of a device, differently from the conventional apparatus disclosed in Japanese Patent Application Non-examined Publication No. H5-341798.   Time for a recognition process after the first utterance can therefore be reduced.

In the present invention, advantageously, the simplified and adequate speaker adaptation using less utterance has a better speech recognition performance than a conventional adaptation.   The conventional adaptation is, for example, a speaker adaptation by only selection of one from a plurality of trained patterns by characteristic or by only distortion coefficient of spectral frequency of an input speech sound.   The speaker adaptation in coincidence with the present invention can reduce a number of speaker trained patterns in combination with a distortion coefficient of the spectral frequency of the input speech sound.   The speaker adaptation can also provide an advantage of reducing a memory capacity.